# SWAN/SIOC: Aligning Scientific Discourse Representation and Social Semantics

Alexandre Passant[1], Paolo Ciccarese[2, 3], John G. Breslin[1, 4], Tim Clark[2, 3]

[1] Digital Enterprise Research Institute, National University of Ireland, Galway,
IDA Business Park, Lower Dangan, Galway, Ireland
{firstname.secondname}@deri.org
[2] Massachusetts General Hospital, Boston, MA 02129 USA
[3] Harvard Medical School, Boston, MA 02115 USA
paolo.ciccarese@gmail.com, twclark@nmr.mgh.harvard.edu
[4] School of Engineering and Informatics, National University of Ireland, Galway,
University Road, Galway, Ireland
john.breslin@nuigalway.ie

**Abstract.** SWAN/SIOC is an alignment of two Web ontologies that, taken together, represent Scientific Discourse in online communities at different levels of granularity (content items and discourse elements). The goal of this alignment is to make the discourse structure and component relationships much more accessible to computation, so that information can be navigated, compared and understood in a context far better than is currently possible, both across and within domains. This paper describes these two models and their alignment to support research in Health Care and Life Sciences, as well as an overview of projected future work on the topic.

**Keywords:** Scientific Discourse Representation, HCLS, Social Semantic Web, Ontology Alignment, SWAN, SIOC.

## 1 Introduction

Semantic Web technologies allow us to provide interoperable and structured data on the Web, enabling a paradigm shift from the current Web of Documents towards a Web of Data[1]. An increasing number of Semantic Web applications have been deployed in various environments and one of the most popular examples is related to the Social Web context, or what is termed "Web 2.0" [1], [2]. This field is also known as the Social Semantic Web, where social aspects (such as data sharing, tagging, etc.) are combined with formal and structured representations in order to provide human- and machine-readable content. Among the various vocabularies developed in this area, a leading example is SIOC (Semantically-Interlinked Online Communities) [3].

Moreover, various research efforts have been carried out on representing argumentative discussions and scientific discourse using Semantic Web technologies.

---

[1] http://www.w3.org/2001/sw/

A working example of the latter is represented by the SWAN (Semantic Web Applications in Neuromedicine) project [4]. SWAN aims to develop a practical, common, semantically-structured framework for scientific discourse that has initially been applied to (but is not limited to) significant problems in Alzheimer Disease (AD) research.

However, so far, there has not been much joint work involving the Scientific Discourse Representation and Social Semantic Web communities, while there are obviously strong ties between both, as scientific argumentation often happens within communities of interest, on online platforms such as blogs, wikis, or in online scientific publishing.

In this paper, we present the SWAN/SIOC project that aims to bridge the gap between Scientific Discourse Representation and the Social Semantic Web, by defining a coherent ontology capable of representing both high-level descriptions of communities (thanks to SIOC) and argumentative discussions (using SWAN).

In the next section, we will introduce both the SWAN and SIOC ontologies. Then, we will describe the various alignments that we have defined between both, in terms of new classes and mappings between classes and properties from these two models, leading to the SWAN/SIOC ontology. We will also present one example of data querying focusing on the relevance of such an alignment. Finally, we will present related work on the topic before concluding the paper with an overview of future work.

## 2 Overview of SWAN and SIOC

We provide an overview of SWAN and SIOC in this section, with motivating use cases. We will focus especially on their relevant features in the context of the SWAN/SIOC integration, described in the following section, which is targeted for use within the Health Care and Life Sciences domain.

### 2.1 SWAN: Semantic Web Applications for Neuromedicine

The SWAN project[2] attempts to model scientific discourse about Alzheimer disease and its supporting evidence in a rich and extensible way that is compatible with the way the domain of Alzheimer Disease (AD) research functions as a technology-mediated knowledge ecosystem. The SWAN knowledge base, for which the SWAN ontology functions as a schema, consists of a semantically-structured network of hypotheses, claims, dialogue, evidence, publications and digital repositories, incorporating and extending such knowledge. Curators of the SWAN knowledge base have catalogued and annotated dozens of etiopathological models of AD, in collaboration with many of the leading researchers in the field. Interestingly, SWAN can not only show the evidentiary support (if any) for each claim in such models, but also a claim's relationships (support, conflict, alternative interpretation,

---

[2] http://swan.mindinformatics.org

neutral) with claims in other models. AD researchers can access the knowledge base online and they can use it orient themselves to new discoveries in the field and how they are related to current models, and to discuss new claims in the literature.

The SWAN ontology[3] was created and continues to evolve in the context of building actual applications for biomedical researchers, as well as through extensive discussions and collaborations within the larger bio-ontologies community, including the NeuroCommons effort [5], the Neuroscience Information Framework [6], [7], and Protein Ontology projects [8]. The SWAN ontology ecosystem consists of a set of modules each covering a specific topic (Figure 1). Three of these modules are of particular interest for the SWAN/SIOC project:

– the Scientific Discourse Relationships module[4], which collects some of the relationships used for modeling the discourse, such as `swandisrel:agreesWith`;
– the Scientific Discourse module[5], which provides a set of classes and properties to represent discourse elements, such as `swanscidis:DiscourseElement` or `swanscidis:ResearchQuestion`; and
– the Citations module[6] , which aims to model the various citation elements (such as `swanscit:Citation` or `swanscit:JournalArticle`) that occur in scientific publishing.
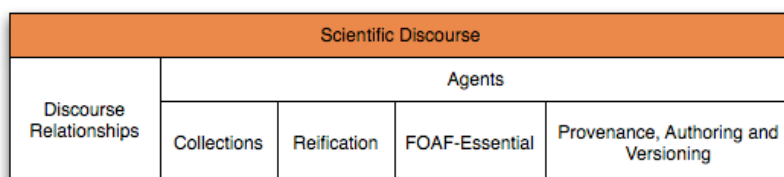
| Scientific Discourse | | | | |
|---|---|---|---|---|
| Discourse Relationships | Agents | | | |
| | Collections | Reification | FOAF-Essential | Provenance, Authoring and Versioning |

**Fig. 1.** Modules in the SWAN ontology

## 2.2 SIOC: Semantically-Interlinked Online Communities

In the Health Care and Life Sciences domain, many researchers are now using Web 2.0 tools or services to share their knowledge in addition to providing traditional publications (research papers). For example, scientists and researchers use blogs to post about their experiments or recent publications that they have read; they use wikis to build information collaboratively (from encyclopedias to project proposals); and they may even participate in scientific social networks, such as Nature Networks.

However, while these services help in the process of publishing information, they generally function as independent and isolated data silos. Therefore, it is difficult to retrieve and to browse information spread across various platforms. A researcher

---

[3] http://swan.mindinformatics.org/ontology.html

[4] http://swan.mindinformatics.org/spec/1.2/discourserelationships.html

[5] http://swan.mindinformatics.org/spec/1.2/scientificdiscourse.html

[6] http://swan.mindinformatics.org/spec/1.2/collections.html

interested, for example, in AD will have to discover and browse various services on his or her own to find relevant information (if any exists).

The aim of the SIOC project [1] is to solve such issues by providing interoperability between these applications using Semantic Web technologies, through an ontology and a set of related tools. In the context of this paper, SIOC can provide improved knowledge sharing and retrieval in scientific communities using these services. In particular, the SIOC Core ontology[7] defines a set of core classes and properties to represent these communities (see Figure 2), while the SIOC Types[8] module provides a more fine-grained set of classes to define content types posted on the Web (such as differentiating a blog entry from a wiki page via the `sioct:BlogPost` and `sioct:WikiArticle` classes).
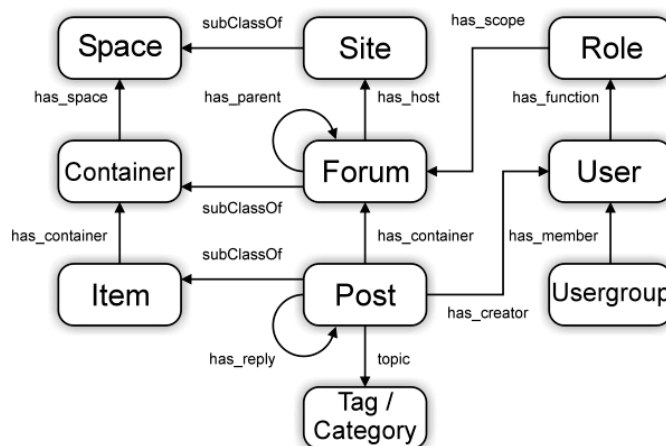


**Fig. 2.** The SIOC Core ontology

For example, imagine that ACME Pharma uses various blogs, wikis and microblogging applications to enable communication and knowledge sharing between its different research teams. By providing SIOC exports of all this data, and through the use of existing applications, APIs and a central RDF repository to store this data, it is then possible to query it from a single place using uniform SPARQL queries. Moreover, these queries can take advantage of the SIOC Types module, for example, so as to retrieve only instances of `sioct:WikiArticle` or `sioct:BlogPost` depending on the requested sources of information.

## 3   SWAN/SIOC: Aligning SWAN and SIOC

As described in the previous sections, SWAN and SIOC function in a

---

[7] http://rdfs.org/sioc/spec
[8] http://rdfs.org/sioc/types

complementary way: SWAN provides fine-grained modeling primitives for scientific discourse elements while SIOC can represent the more generic contributions found in online communities. Bridging both would therefore help one to browse these communities and their related discussions using various levels of granularity, e.g. at the item level (thanks to SIOC) and then zooming in to the discussion level (using SWAN). For example, considering the previous ACME use case, the items could be connected to each other using SIOC (related posts, replies, etc.), but also the kind of relationship that they have to each other could be specified using SWAN (agreement, disagreement, supporting hypothesis, etc.). Then, users would be able to browse information from the various ACME Social Web applications using different layers, depending on their query and the kind of information they want to retrieve. Moreover, combining these two levels also provides advanced querying patterns. When browsing a wiki (represented using SIOC), one could identify all elements that support or contradict the claims of that wiki page (using SWAN) and then filter by content types, i.e. blog posts (using SIOC).
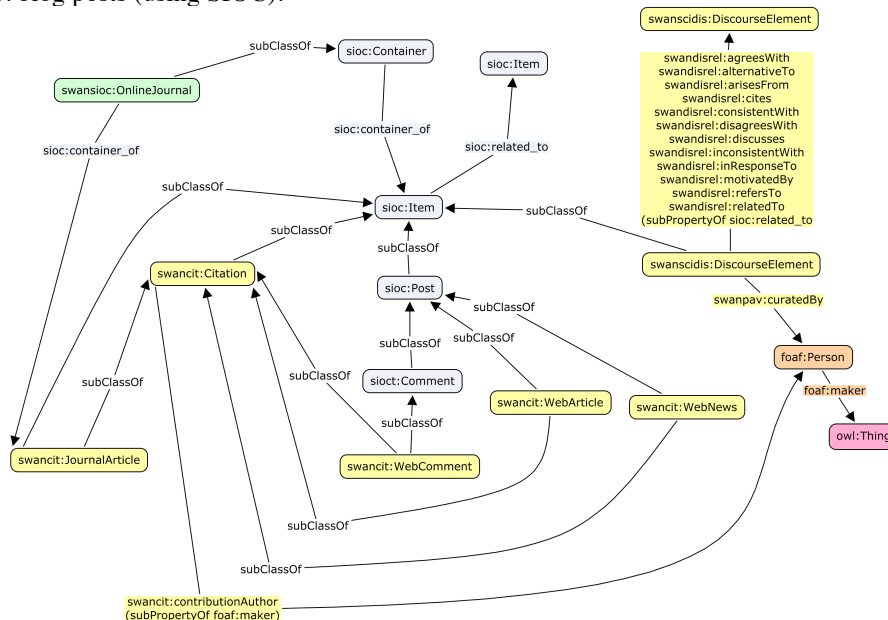


**Fig. 3.** Overview of the SWAN/SIOC alignments

In order to bridge the SWAN and SIOC ontologies, alignments between these two models have been provided, as we will now detail. These different mappings have been defined in a SIOC module available at http://rdfs.org/sioc/swan, an overview of which is given in Figure 3. This module imports the SIOC Core Ontology and its Types module, as well as the SWAN Ontology, via its OWL definition file[9]. It has been validated as OWL-DL (using Pellet[10] version 1 [9], [5]), with a *SHIF*(D) expressivity.

---

[9] http://swan.mindinformatics.org/ontologies/1.2/swan.owl
[10] http://clarkparsia.com/pellet/

### 3.1 Adapting the SIOC Ontology to OWL-DL

Previously, the SIOC Core ontology was designed in RDFS, whilst also being an OWL-Full ontology. However, one of the requirements for the SWAN project and related services is to be able to reason on SWAN data to, for example, use OWL cardinality constraints defined in the Scientific Discourse module to verify that each instance of `swanscidis:DiscourseElement` has at maximum one `swanscidis:title`. Using the SIOC Ontology with SWAN would not ensure that such reasoning could be achieved in a finite time, because of the OWL-Fullness of SIOC. Therefore, and as we needed the computability of OWL-DL, we adapted the existing RDFS SIOC Core Ontology to OWL-DL by:

– Declaring the value of `rdf:type` as being `owl:Class` for some classes defined in external ontologies and used in the SIOC Core Ontology, such as `foaf:Person`, since we do not use `owl:imports` to include these external ontologies in SIOC but require that typing to make the ontology OWL-DL;
– Adapting some disjointness statements in the SIOC ontology to make them compliant with OWL-DL axioms, using `owl:disjointWith` properties.

### 3.2 Class Mappings

In addition to the aforementioned changes to the SIOC Core ontology, various classes from the SWAN ontology have been mapped to classes in the SIOC Core ontology. From SWAN Scientific Discourse, the following classes have been defined as subclasses (via `rdfs:subClassOf`) of sioc:Item:

– `swanscidis:DiscourseElement`;
– `swanscidis:ResearchStatement`;
– `swanscidis:ResearchQuestion`;
– `swanscidis:ResearchComment`.

In addition, from SWAN Citations, the following mappings have been defined:

– `swancit:Citation` and `swancit:JournalArticle` are subclasses of `sioc:item`;
– `swancit:WebArticle` and `swancit:WebNews` are subclasses of `sioc:Post`;
– `swancit:WebComment` are subclass of `sioc:Comment`.

Consequently, most of the SWAN elements became subclasses of the `sioc:Item` class, since `sioc:Post` is also defined as a subclass of that resource.

However, as one can see when observing these mappings, some of them are redundant. For example, we explicitly assert that `swancit:JournalArticle` is a subclass of `sioc:item`, though this could be inferred from the assertions that `swancit:JournalArticle` is a subclass of `swancit:Citation` and `swancit:Citation` is in turn a subclass of `sioc:Item`.

In addition, a new class has been introduced in the SWAN/SIOC module for online journals (these are websites where immutable articles are published and comments are allowed on them). `swansioc:OnlineJournal` is defined as a subclass of `sioc:Container`, and can be used to represent online publication venues such as

StemBook[11].

Finally, there may be a need to state that a particular `swanscidis:DiscourseElement` is a part of a `sioc:Item`, for example, to represent that a particular hypothesis is part of a blog post, and then to identify in which forums this blog post is contained. This item-to-item inclusion is not specific to the SWAN use case and can already be achieved thanks to the `dcterms:hasPart` property from Dublin Core, as suggested in the SIOC specification document.

### 3.3 Property Mappings

In addition to the previous classes, mappings have been defined between various properties of the SWAN Scientific Discourse Relationship and the `sioc:related_to` property of the SIOC Core ontology. The following properties use this mapping, and this permits us to infer that two items are related to each other as soon as there is a particular discourse relationship between both:
– `swandisrel:agreesWith`;
– `swandisrel:alternativeTo`;
– `swandisrel:arisesFrom`;
– `swandisrel:cites`;
– `swandisrel:consistentWith`;
– `swandisrel:disagreesWith`;
– `swandisrel:discusses`;
– `swandisrel:inconsistentWith`;
– `swandisrel:inResponseTo`;
– `swandisrel:motivatedBy`;
– `swandisrel:refersTo`;
– `swandisrel:relatedTo`.
Once again, some of these mappings may be redundant, since they can inherit from the `swandisrel:relatedTo` property, but we provide these for the same reasons as specified earlier for the class mappings.

## 4  Querying Data Using the SWAN/SIOC Alignments

In order to give an overview of the advantages achieved using these alignments, we ran an initial experiment by querying SWAN data using SPARQL queries based on SIOC, hence benefiting from the various mappings between classes and properties that we have already described.

We generated a set of N random instances of `swanscidis:DiscourseElement`, linked to each other using each of the 13 relationships in the Scientific Discourse Module, hence providing a dataset of $N+13*N*(N-1)$ triples[12]. Then, we ran a simple SPARQL query using SIOC patterns, identifying all distinct couples of related items within the dataset (this kind of query

---

[11] http://www.stembook.org/
[12] N triples for instances generation and 13*N*(N-1) for the relationships

often being used in SIOC applications to identify related posts on the Web):

```
PREFIX sioc: <http://rdfs.org/sioc/ns#>
SELECT DISTINCT ?s ?o
WHERE {
  ?s sioc:related_to ?o .
  ?s a sioc:Item . ?o a sioc:Item .
}
```

The query was run using Pellet 2 (making use of its OWL-DL SPARQL capabilities) on a 2.53 MHz MacBook Pro with 4 GB RAM. As expected, we retrieved a list of N(N-1) answers each time, hence being able to simply express queries over SWAN data using SIOC patterns.

In addition, we tried each query using both the full property mappings and with a single mapping between `swandisrel:related_to` and `sioc:related_to` in order to evaluate the influence of our choice of mappings' redundancy over computation time, as we expressed previously. The results for various values of N are described below (times are given in milliseconds). As one can see, while the full mappings are not a good choice when dealing with a small number of statements, it becomes interesting when the number of statements grows. Hence, since SWAN knowledge bases generally contain millions of triples, we believe our choice was accurate and enables faster computation of SPARQL queries using SIOC patterns over SWAN data.

**Table 1.** Computation time (in msec.) for SPARQL queries using the SWAN/SIOC mappings.

| N | Triples | Time with full mappings | Time with single mapping |
|---|---|---|---|
| 2 | 2 | 9885 | **8469** |
| 5 | 265 | 8426 | **8254** |
| 10 | 1180 | **8338** | 8502 |
| 50 | 31900 | **17471** | 15441 |
| 100 | 128800 | **40640** | 45178 |
| 200 | 517600 | **188655** | 195558 |
| 300 | 1166400 | **418566** | 462990 |

## 5  Related Work

Related work includes IBIS [13] and gIBIS [14], or (graphical) issue-based information systems, which use argumentative discussions in the process of solving design and planning issues and provide detailed models for links between conversations.

An argumentation module extension to SIOC has been provided to allow one to formulate agreement and disagreement between SIOC content items[13] [10]. The properties and classes defined in this module can then be related to other

---

[13] http://rdfs.org/sioc/arguments

argumentation models such as SALT[14] (Semantically Annotated LaTeX) [11] and IBIS. Some reply types such as agree or disagree have also been ontologised by the W3C[15].

Another recent effort that may align well with the SWAN/SIOC project is aTags [12], which combines discourse representation and paradigms of the Social Web by providing a way to create statements (claims or hypothesis) using free tagging combined with knowledge bases such as DBpedia.

# 6  Conclusion

In this paper, we introduced the motivations for the SWAN/SIOC initiative and detailed the mappings that have been created between the SWAN and SIOC ontologies in order to enable better computation and understanding of Scientific Discourse in online communities. We also demonstrated how these mappings could be used for data querying in order to provide both high-level and more fine-grained descriptions of relations between statements.

Future work will consist of building applications on the top of these new alignments, especially within the Science Collaboration Framework[16]. In addition, we will also investigate how a similar process of mappings could be applied to other ontologies relevant to Scientific Discourse Representation, hence providing a complete and integrated framework for machine-readable discourse in online scientific communities.

We hope that SWAN/SIOC will be a first step towards a more comprehensive work on aligning different frameworks for discourse representation in online communities.

# Acknowledgements

---

[14] http://salt.semanticauthoring.org/
[15] http://www.w3.org/2001/12/replyType
[16] http://sciencecollaboration.org/

# References

[1] O'Reilly T. What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. In: O'Reilly Network; 2005.

[2] Breslin J.G, Bojars U, Passant A, Fernandez S, Decker S. SIOC: Content Exchange and Semantic Interoperability Between Social Networks. In: W3C Workshop on the Future of Social Networking. Barcelona, Spain; 2009.

[3] Breslin J.G, Harth A, Bojars U, Decker S. Toward Semantically-Interlinked Online Communities. Lecture Notes in Computer Science l2005;3532/2005: 500-514.

[4] Ciccarese P, Wu E, Wong G, Ocana M, Kinoshita J, Ruttenberg A, Clark T. The SWAN biomedical discourse ontology. J Biomed Inform l2008;41: 739-51.

[5] Ruttenberg A, Rees JA, Samwald M, Marshall MS. Life sciences on the Semantic Web: the Neurocommons and beyond. Brief Bioinform l2009: bbp004.

[6] Gardner D, Akil H, Ascoli GA, Bowden DM, Bug W, Donohue DE, Goldberg DH, Grafstein B, Grethe JS, Gupta A, Halavi M, Kennedy DN, Marenco L, Martone ME, Miller PL, Muller HM, Robert A, Shepherd GM, Sternberg PW, Van Essen DC, Williams RW. The neuroscience information framework: a data and knowledge environment for neuroscience. Neuroinformatics l2008;6: 149-60.

[7] Gupta A, Bug W, Marenco L, Qian X, Condit C, Rangarajan A, Muller HM, Miller PL, Sanders B, Grethe JS, Astakhov V, Shepherd G, Sternberg PW, Martone ME. Federated access to heterogeneous information resources in the Neuroscience Information Framework (NIF). Neuroinformatics l2008;6: 205-17.

[8] Natale DA, Arighi CN, Barker WC, Blake J, Chang TC, Hu Z, Liu H, Smith B, Wu CH. Framework for a protein ontology. BMC Bioinformatics l2007;8 Suppl 9: S1.

[9] Sirin E, Parsia B, Grau BC, Kalyanpur A, Katz Y. Pellet: A practical OWL-DL reasoner. Journal of Web Semantics l2007;5: 51-53.

[10] Lange C, Bojars U, Grosza T, Breslin JG, Handschuh S. Expressing Argumentative Discussions in Social Media Sites. In: 1st International Workshop on Social Data on the Web (SDOW 2008) at the 7th International Semantic Web Conference (ISWC 2008). Karlsruhe, Germany; 2008.

[11] Groza T, Handschuh S, Moller K, Decker S. SALT - Semantically Annotated LaTeX for Scientific Publications. In: The Semantic Web: Research and Applications. Berlin / Heidelberg: Springer; 2007, p. 518-532.

[12] Samwald M, Stenzhorn H. Simple, ontology-based representation of biomedical statements through fine-granular entity tagging and new web standards. In: Bio-Ontologies 2009 - Knowledge in Biology. Stockholm, Sweden; 2009.

[13] W. Kunz and H.W. J. Rittel. Issues as Elements of Information Systems. Technical Report WP-131, University of California, Berkeley, 1970.

[14] J. Conklin and M. Begeman. gIBIS - A Hypertext Tool for Exploratory Policy Discussion. In The Conference on Computer-Supported Cooperative Work, Proceedings, pages 140–152, 1988.